

Calibración de modelos

Gerardo Martín

Introducción

Los Modelos

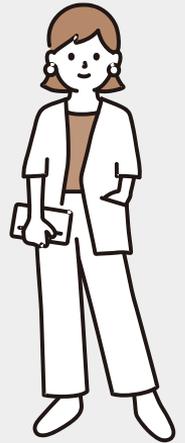
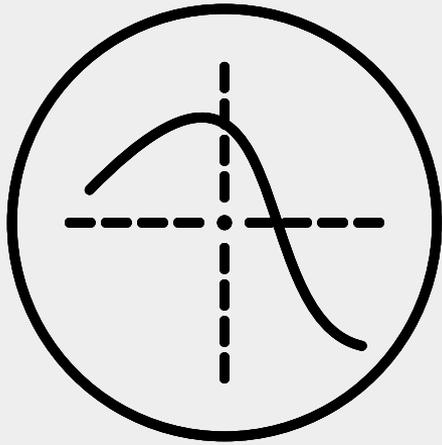
¿Qué son?

Tipos

Filosofías

Ejemplos

¿Qué son?



¿Qué son?

Simplificaciones, deben
representar algo en suficiente
detalle para reconocerlo....

¿Qué son?

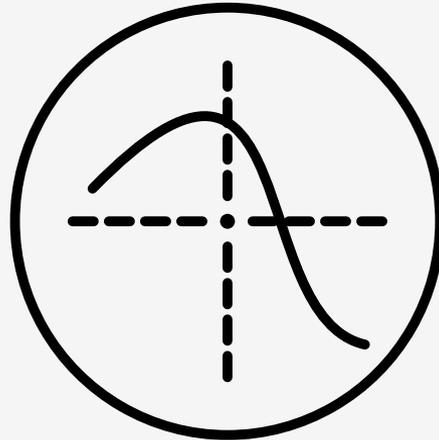
Los modelos que nos
concernen en ecología ...

Representaciones matemáticas
de los fenómenos ...

Tipos

Mecanismos

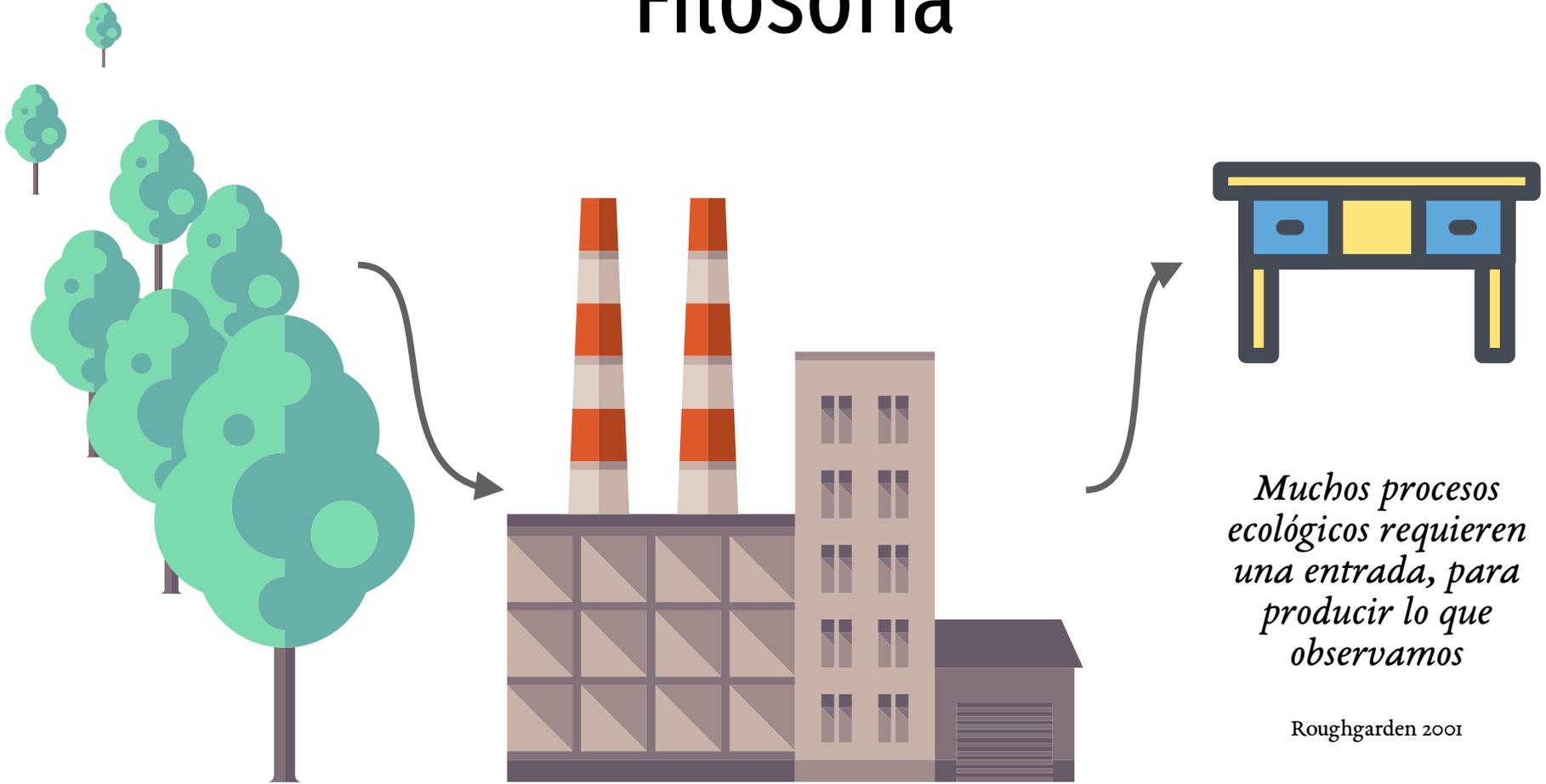
Basados en procesos
Mecanísticos



Empíricos
Estadísticos
Correlativos

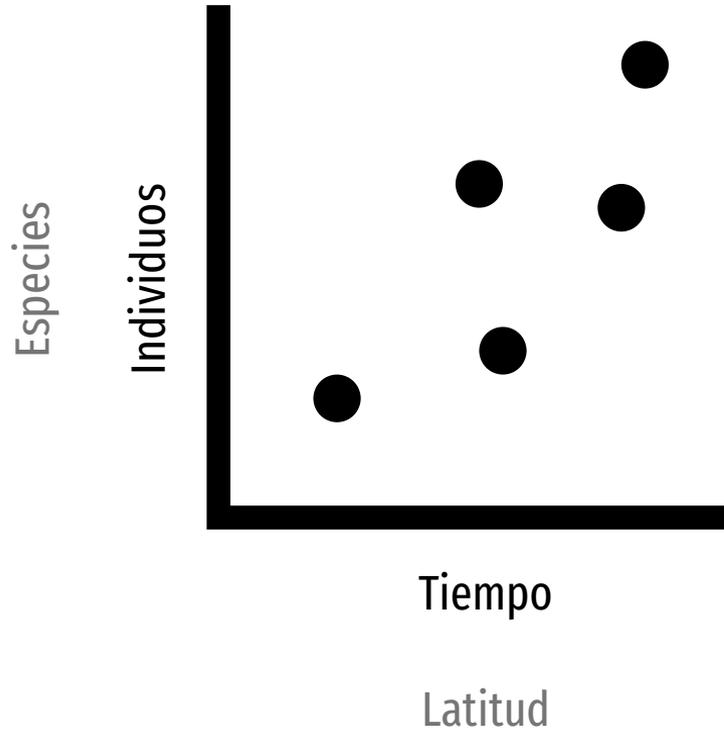
Patrones

Filosofía



Roughgarden 2001

Filosofía



Generalmente
observamos un
patrón

¿Qué lo generó?

H_0 : Más nacimientos
que muertes

H_1 : El tiempo

$$\text{Ind} = \text{Nacimientos} - \text{Muertes}$$

$$\text{Ind} = \beta t$$

$\text{Ind} = \text{Nacimientos} - \text{Muertes}$

$$\text{Ind} = \beta t$$

¿Cuál hipótesis es biológicamente más robusta?

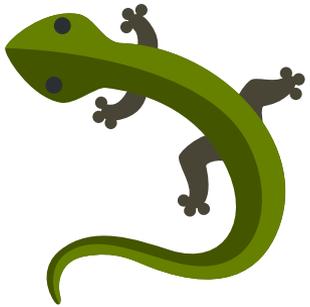
¿Cuál hipótesis es más fácil de probar?

¿Con cuál hipótesis aprendemos más de nuestro sistema?

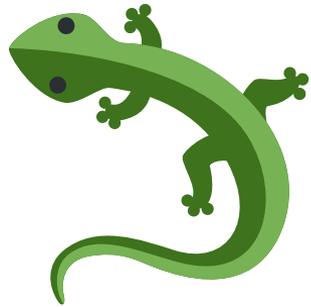
Expresión matemática representa mecanismo → Modelo
mecanístico/basado en proceso

Expresión matemática describe el patrón →
Empírico/correlativo

Ejemplos en MNE

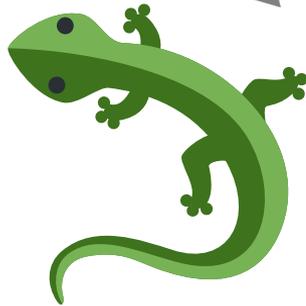
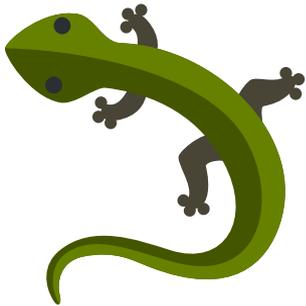
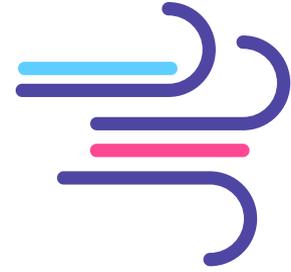
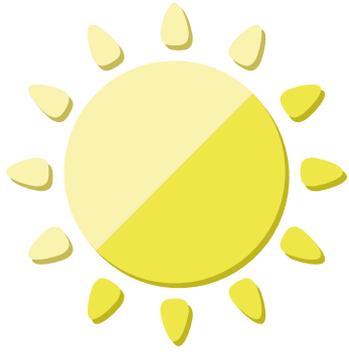


VS



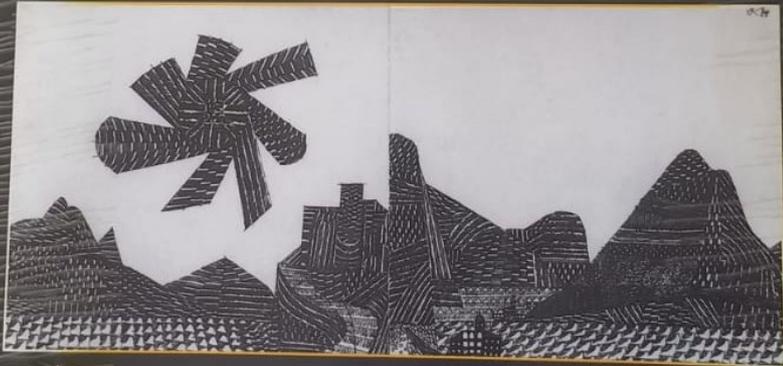
VS





Hollberton

PRIMER OF ECOLOGICAL THEORY



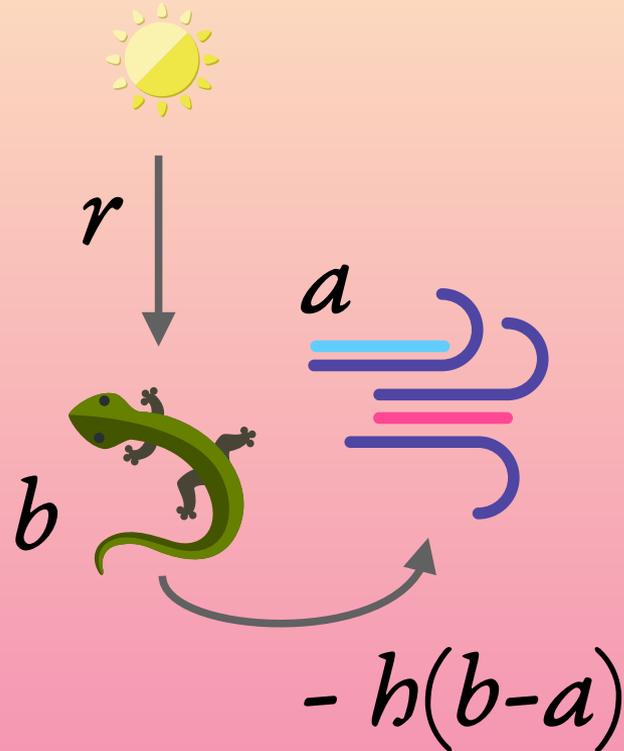
Jonathan Roughgarden

“Getting
comfortable
outdoors”

Determinación de la temperatura corporal de equilibrio de un organismo ectotermo pequeño

¿Qué determina la temperatura corporal?

Termodinámica: objetos absorben y emiten calor



$$f = r - h(b - a)$$

f – Flujo de calor (cal/hr)

r – Radiación solar (~ 1000 cal/hr)

h – Coeficiente de transferencia
convectiva de calor (cal/hr $^{\circ}\text{C}$)

b – temperatura corporal ($^{\circ}\text{C}$)

a – temperatura del aire ($^{\circ}\text{C}$)

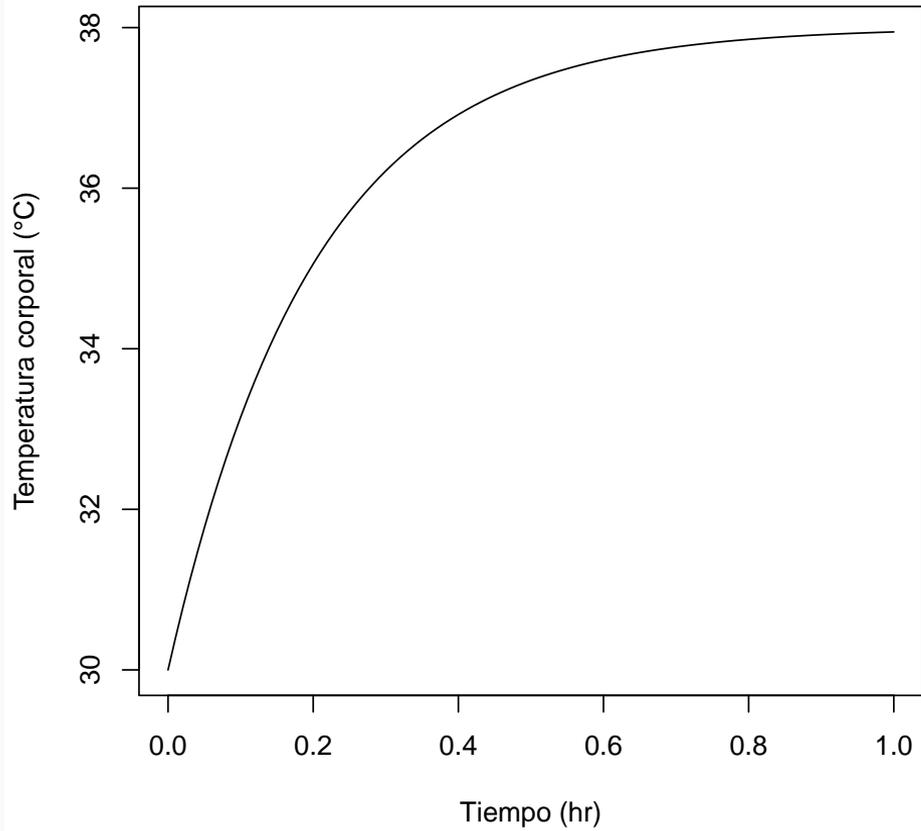
f , el flujo de calor, = cambio temporal de temperatura, y añadiendo el peso w del organismo:

$$\frac{dc}{dt} = \frac{r - h(b - a)}{w}$$

dc/dt es integrable, para obtener una función donde
sustituir las unidades de tiempo:

$$\int \frac{db}{dt} = b(t) = \frac{r}{h} + a + e^{-\frac{h}{w}t} \left(-\frac{r}{h} - a + c \right)$$

c es la constante de integración, y representa la temperatura corporal inicial b_0 .



$$r = 1000 \text{ cal/hr)}$$

$$h = 50 \text{ cal/hr } ^\circ\text{C}$$

$$b_o = 30 \text{ } ^\circ\text{C}$$

$$a = 18 \text{ } ^\circ\text{C}$$

¿Y ahora?

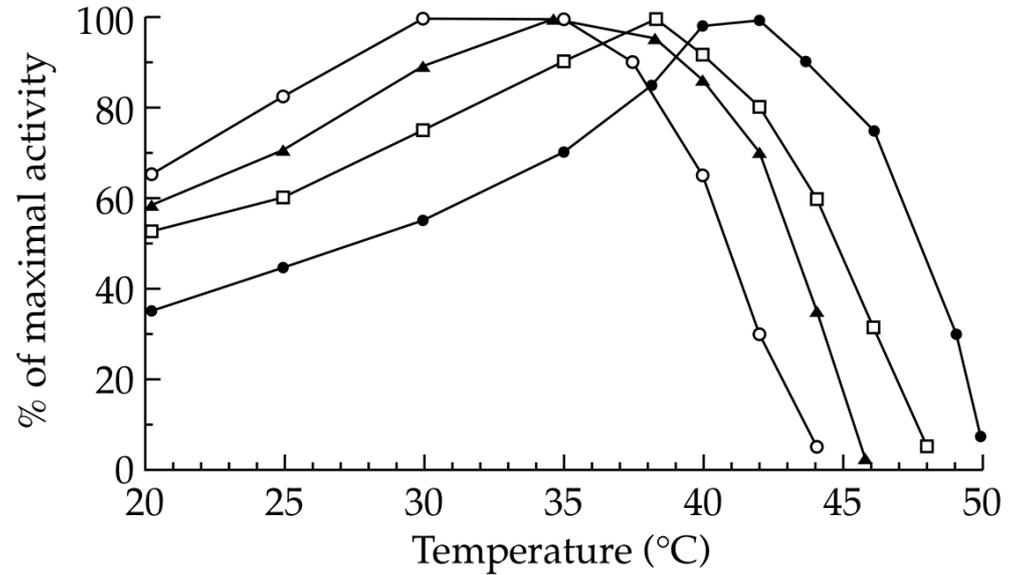
- Biológicamente, ¿de qué nos sirve conocer la temperatura corporal de equilibrio?
- ¿Tiene alguna utilidad aplicada?

Thermal Adaptation

A Theoretical and Empirical Synthesis



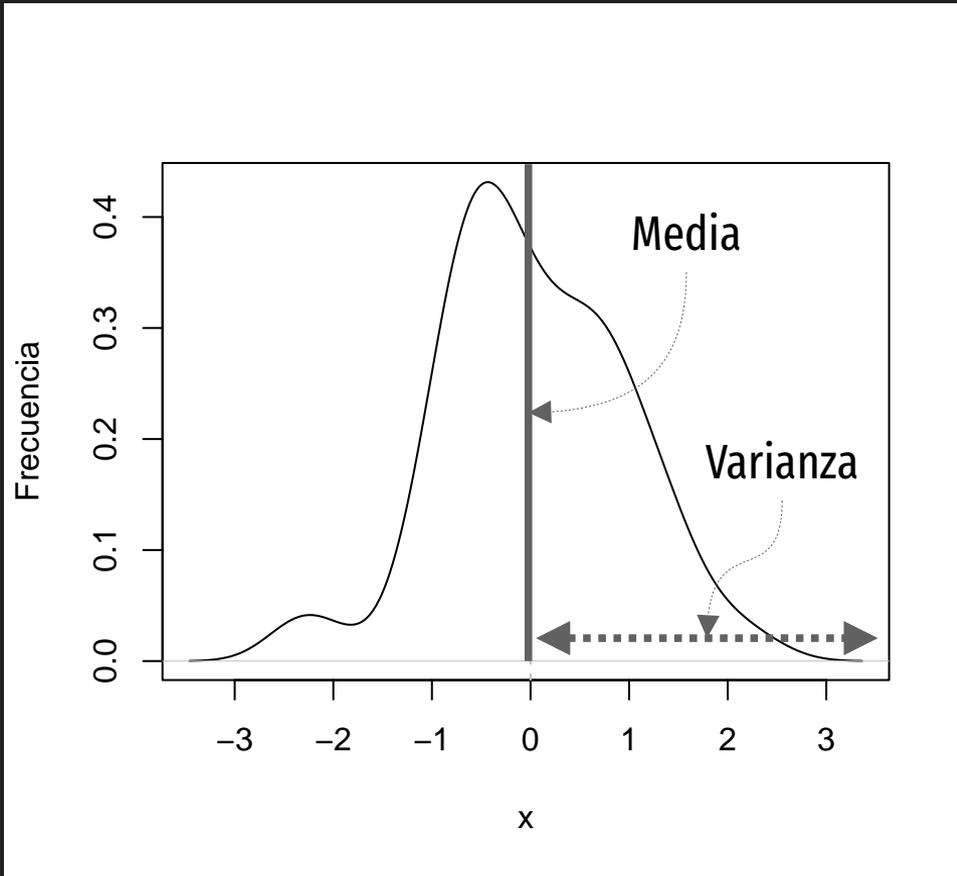
Michael J. Angilletta Jr.



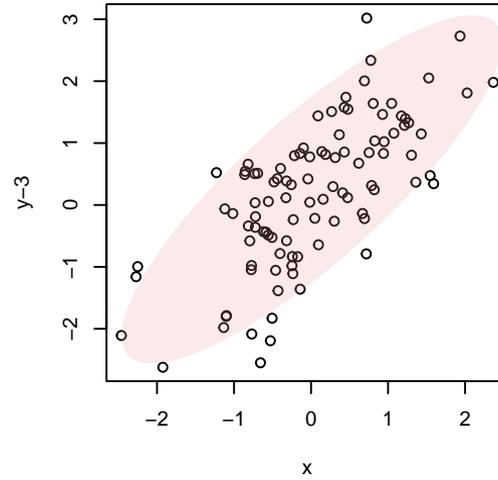
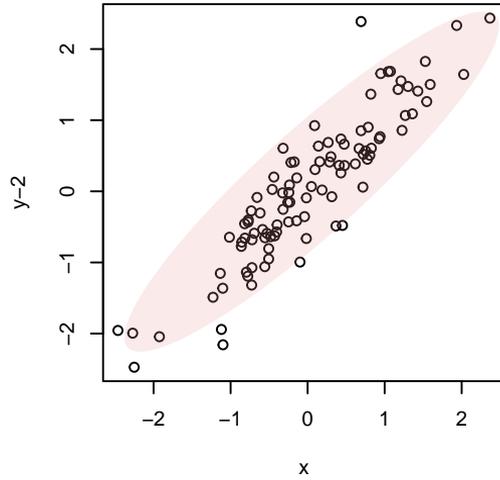
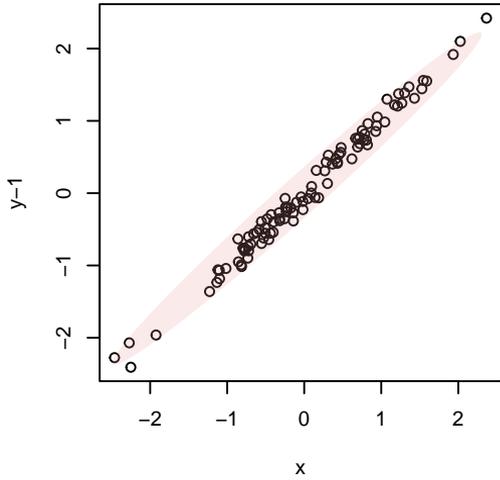
Actividad de ATPasa en función de la temperatura corporal en tres spp. de lagartijas

Consideraciones básicas para la modelación correlativa

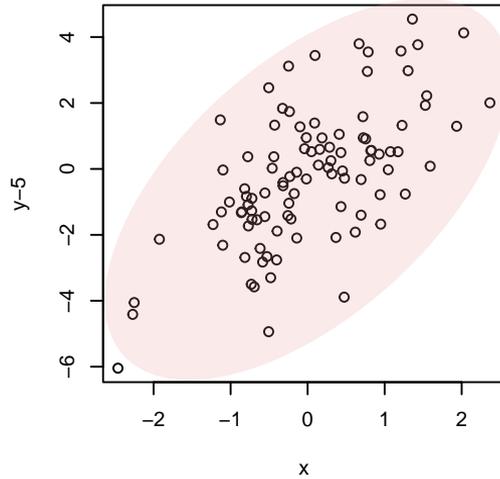
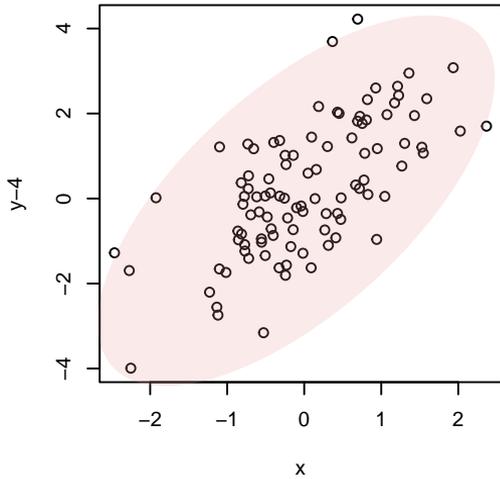
¿Por qué los llamamos “modelos correlativos”?



Mucha varianza de y explicada por x



Poca varianza explicada



Con los modelos estadísticos
queremos encontrar qué
explica la varianza de nuestros
datos de respuesta

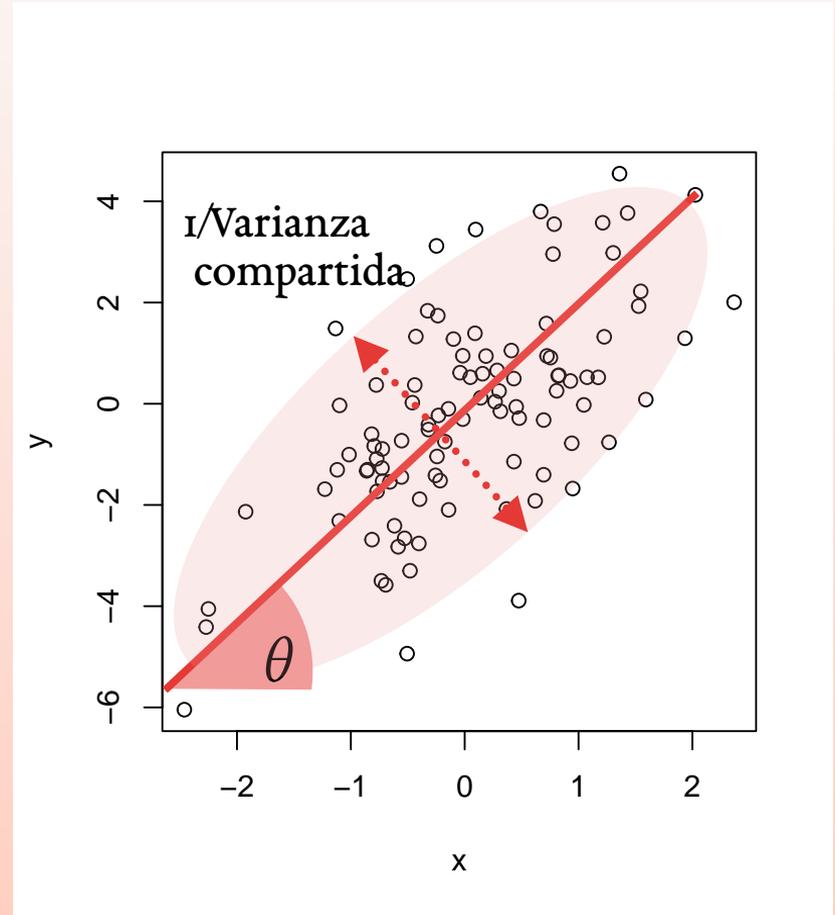
Correlación de Pearson ~ Coeficiente de determinación

$$r = \text{signo}(\beta) \sqrt{R^2}$$

$$y = a + \beta x$$

$$\beta = \tan \theta$$

Regresión lineal, es una extensión de la correlación, para conocer constantes de proporcionalidad β



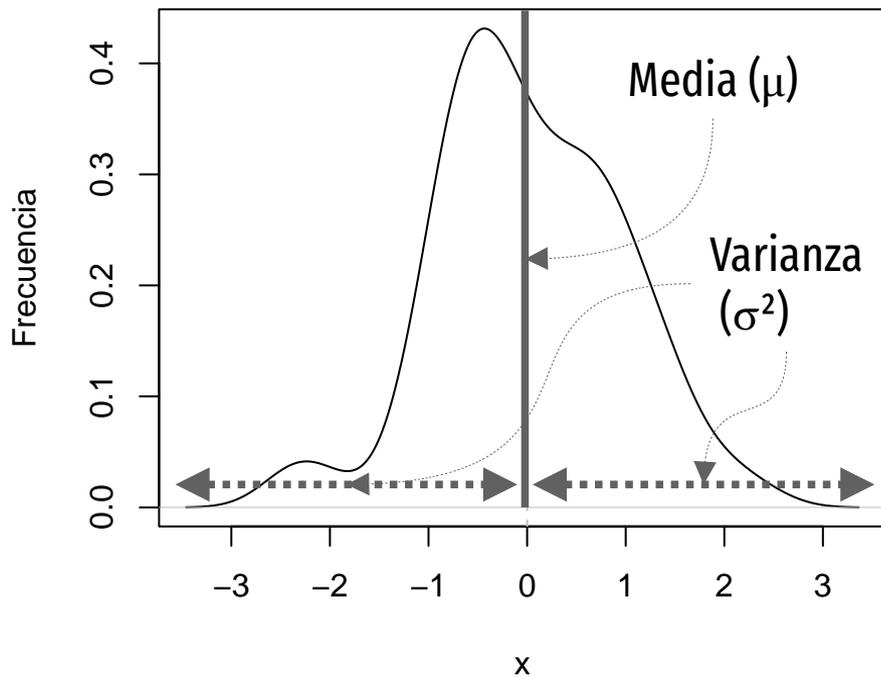
Regresión nos permite construir funciones mediante relaciones empíricas entre variable de respuesta e independiente(s)

Mediante regresión buscamos explicar varianza, similar a correlación

Técnicas avanzadas no establecen relación matemática tan directa como lineal

Tipos de datos y técnicas de regresión

Observación	y_1	y_2	y_3	y_4	y_5	y_6	
1	-0.07	0.31	0	0.37	0	4	2
2	0.09	0.42	5	0.99	1	6	4
3	1.08	0.82	6	0.62	1	5	4
4	-0.17	2.66	3	0.01	0	7	3
...	-0.27	1.34	0	0.21	1	7	4



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mu = \sum \frac{x}{n}$$

$$\sigma^2 = \sum \frac{(x-\mu)^2}{n+1}$$

$$x \sim \text{normal}(\mu, \sigma^2)$$

Tipos de variables más comunes

y_3	y_5	x, y
0	0	98.1, -19.3
5	1	97.1, -18.9
6	1	97.13, -17.3
3	0	99.5, -20.3
0	1	98.3, -21.01

Conteos, binarias, coordenadas

Conteos, binarias, coordenadas

~

Variable 1 + Variable 2 + + f(Variable 1) +

Calibración, encontrar las constantes β para cada variable ó función de variable:

$$N(X_1, X_2, \dots, X_n) = \alpha + \beta_1 X_1 + \dots + \beta_n X_n + \beta'_1 f(X_1) + \dots$$

$$N(X_1, X_2, \dots, X_n) = \alpha + \beta_1 X_1 + \dots + \beta_n X_n + \beta'_1 f(X_1) + \dots$$

$$N(X_1) = \alpha + \beta_1 X_1$$

$$N(X_1, X_2) = \alpha + \beta_1 X_1 + \beta_2 X_2$$

$$N(X_1, X_2) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta'_1 f(X_1) + \dots$$

$$N(\dots) = \dots$$

Calibración
también implica
seleccionar el
modelo que *mejor*
representa los datos.

Modelos estadísticos *ad-hoc*

Modelos lineales
generalizados (GLM)

Poisson
Binomiales/Logísticos

Generalizaciones
de regresión lineal
ordinaria para
otros tipos de
error

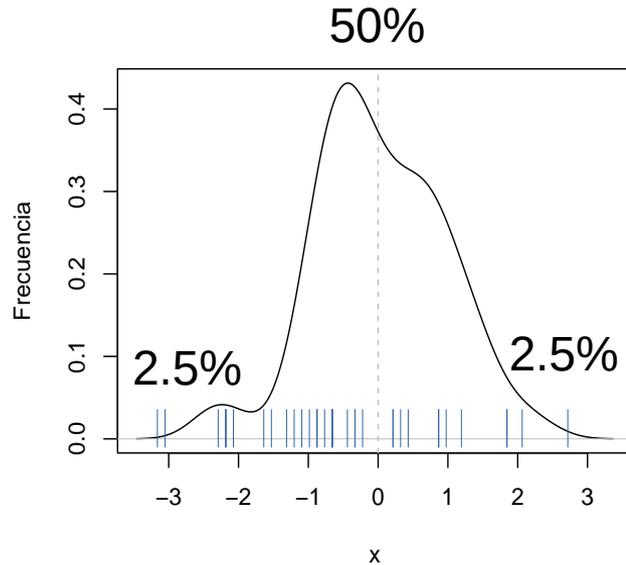
y_3	y_5	x, y
0	0	98.1, -19.3
5	1	97.1, -18.9
6	1	97.13, -17.3
3	0	99.5, -20.3
0	1	98.3, -21.01

Extensiones de GLM,
estadística multivariada

Patrones de puntos,
Maxent
Elipsoides, ENFA, Bioclim

Adaptaciones de
GLMs, definición
del patrón de
puntos

Consideraciones sobre la complejidad del modelo



100 observaciones
 n variables

$$N(X_1, X_2, \dots, X_n)$$

Modelos con más variables
(más complejos) pueden
explicar más varianza de x ,
pero pierden capacidad de
predecir lo que desconocen

¿Cuántas variables necesitamos para explicar todo?

¿Necesitamos explicar todo?

¿Cuántas variables como máximo es aceptable?

¿Hay una complejidad mínima o máxima aceptable?

Consideraciones sobre los datos

x, y
98.1, -19.3
97.1, -18.9
97.13, -17.3
99.5, -20.3
98.3, -21.01

¿Cómo nos hacemos de estas colecciones de coordenadas?

¿Cómo se construye GBIF, Naturalista, etc.?

Short communication

Spatial filtering to reduce sampling bias can improve the performance of ecological niche models



Robert A. Boria^{a,*}, Link E. Olson^b, Steven M. Goodman^{c,d}, Robert P. Anderson^{a,e,f}

^a *Department of Biology, City College of the City University of New York, New York, NY 10031, USA*

^b *University of Alaska Museum, Fairbanks, AK 99708, USA*

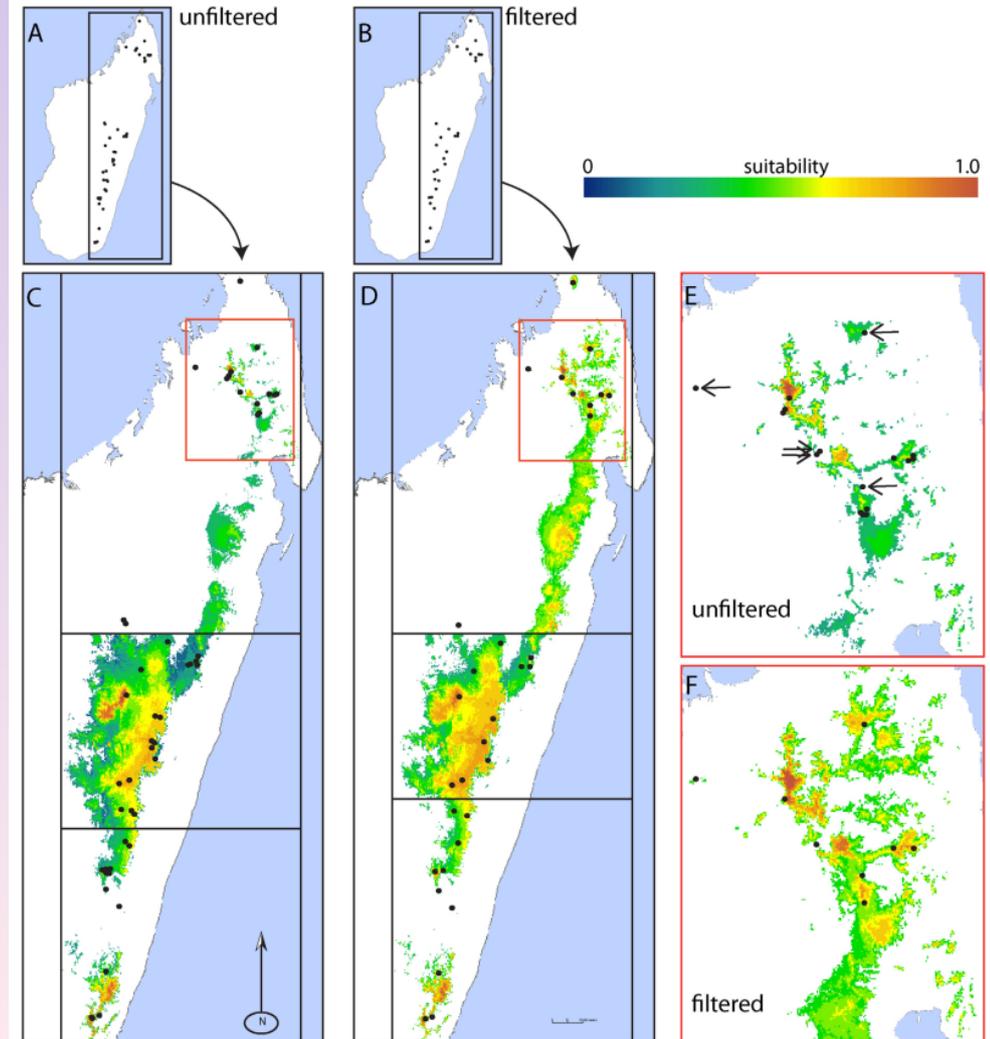
^c *Field Museum of Natural History, Chicago, IL 60605, USA*

^d *Association Vahatra, BP 3972, Antananarivo 101, Madagascar*

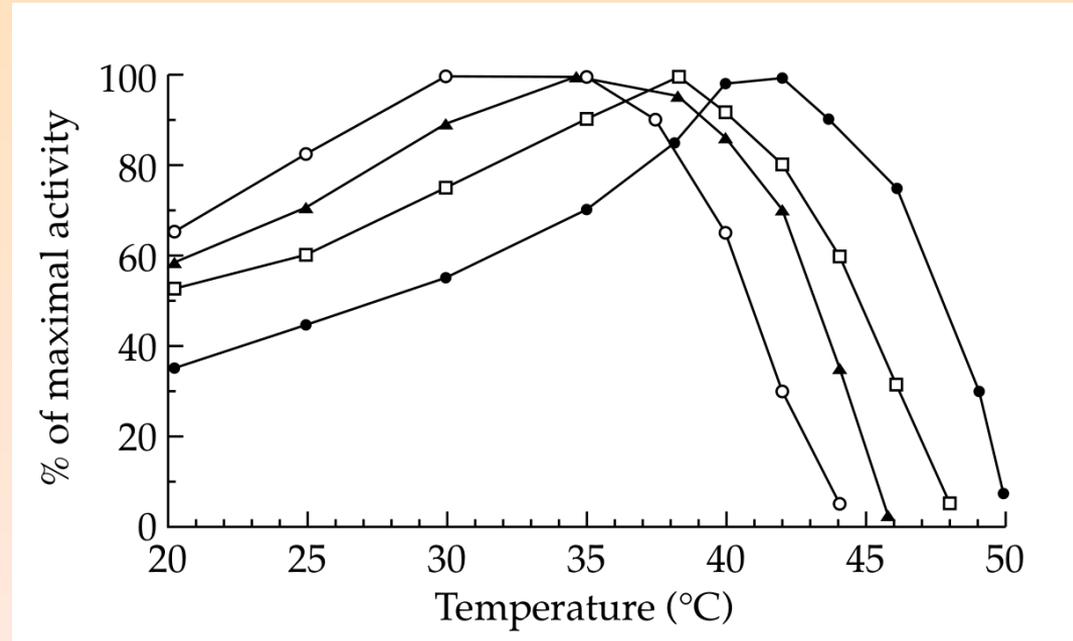
^e *Graduate Center, City University of New York, New York, NY 10016, USA*

^f *Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, New York, NY 10024, USA*

Sobrerrepresentación
ambiental puede generar
modelos
geográficamente sobre-
ajustados, aún cuando el
modelo no sobreajuste
con variables...



Consideraciones sobre los métodos



Efectos ambientales rara vez son lineales

Paréntesis para hablar de los patrones de puntos y sus propiedades... y
sobre el papel de la evaluación en el proceso de calibración